



(19) 대한민국특허청(KR)

(12) 등록특허공보(B1)

(45) 공고일자 2015년12월08일

(11) 등록번호 10-1575010

(24) 등록일자 2015년12월01일

(51) 국제특허분류(Int. Cl.)

G06F 9/455 (2006.01)

(21) 출원번호 10-2013-7009995

(22) 출원일자(국제) 2011년09월20일

심사청구일자 2013년04월19일

(85) 번역문제출일자 2013년04월19일

(65) 공개번호 10-2013-0075777

(43) 공개일자 2013년07월05일

(86) 국제출원번호 PCT/US2011/052413

(87) 국제공개번호 WO 2012/040241

국제공개일자 2012년03월29일

(30) 우선권주장

12/886,138 2010년09월20일 미국(US)

(56) 선행기술조사문헌

US20100223378 A1*

WO2009126154 A1*

US07529822 B2*

JP2005050084 A

*는 심사관에 의하여 인용된 문헌

(73) 특허권자

넷플릭스, 임크.

미국 캘리포니아 로스 가토스 원체스터 씨클 100
(우:95032-7606)

(72) 발명자

오젤, 그레고리 에스.

미국 94114 캘리포니아주 샌 프란시스코 넘버2 수
어드 에스티. 45

이즈라일레브스키, 유리

미국 95070 캘리포니아주 사라토가 사라토가 에이
브이이. 13991

(74) 대리인

장수길, 양영준

전체 청구항 수 : 총 18 항

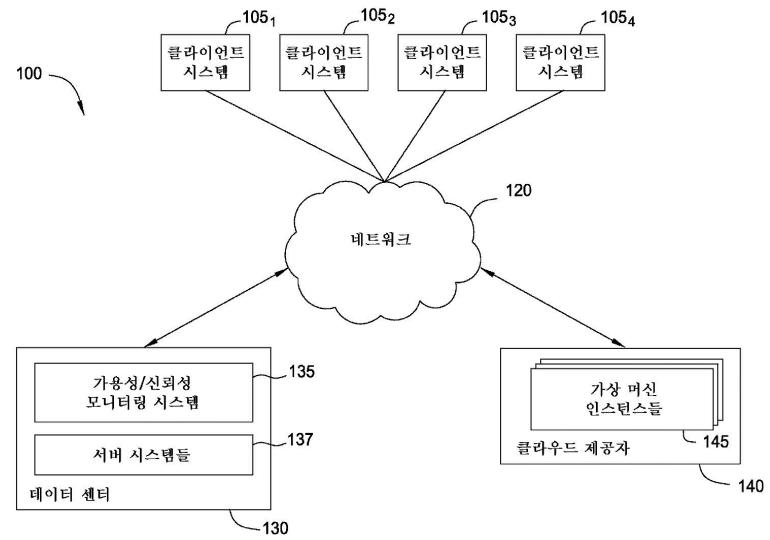
심사관 : 유진태

(54) 발명의 명칭 네트워크화된 애플리케이션들의 리질리언시의 검증

(57) 요약

상호작용하는 서버들의 컬렉션을 사용하여 이용가능해진 네트워크화된 애플리케이션의 리질리언시를 검증하는 기술들이 개시된다. 일 실시예에서, 네트워크 모니터링 애플리케이션은 복특정 간격들에서 각 실행 서버(또는 애플리케이션)를 관찰하고, 하나를 선택하여 종결시킨다. 클라우드 기반 전개의 경우에, 이것은 가장 먼저 인스턴스를 종결시키는 것 또는 서버 상에서 실행하는 프로세스를 종결시키는 것을 포함할 수 있다. 나머지 네트워크화된 애플리케이션에 대한 고장난 서버의 효과들을 관찰함으로써, 제공자는 각 구성요소가 경고 없이 사라지는 임의의 단일 인스턴스를 용인할 수 있는 것을 보장할 수 있다.

대 표 도



명세서

청구범위

청구항 1

네트워크화된 애플리케이션의 리질리언시(resiliency)를 검증하는 컴퓨터 구현 방법으로서, 상기 방법은 처리 유닛에 의해 수행되며, 상기 네트워크화된 애플리케이션을 제공하기 위해 사용되는 네트워크 내의 복수의 활성 애플리케이션 구성요소를 식별하는 단계; 복수의 선택 기준들에 기초하여, 상기 식별된 애플리케이션 구성요소들 중 적어도 하나의 애플리케이션 구성요소를 선택하는 단계; 상기 선택된 활성 애플리케이션 구성요소의 무작위 고장(random failure)을 시뮬레이션하기 위하여 상기 선택된 활성 애플리케이션 구성요소를 종결시키는 단계; 및 상기 선택된 활성 애플리케이션 구성요소의 종결 후에, 상기 네트워크 내의 하나 이상의 나머지 활성 애플리케이션 구성요소들을 모니터링하는 단계를 포함하고, 상기 선택된 활성 애플리케이션 구성요소는 가상 머신 인스턴스 상에서 실행되는 지정된 프로세스를 포함하고, 상기 선택된 애플리케이션 구성요소를 종결시키는 단계는 상기 지정된 프로세스의 실행을 정지시키는 단계를 포함하는 컴퓨터 구현 방법.

청구항 2

제1항에 있어서, 상기 복수의 활성 애플리케이션 구성요소 각각은 컴퓨팅 클라우드에서 실행되는 가상 머신 인스턴스를 포함하는 컴퓨터 구현 방법.

청구항 3

제2항에 있어서, 상기 가상 머신 인스턴스는 오토 스케일링 그룹(auto scaling group) 내에서 유발될 최소 또는 최대 수의 가상 머신 인스턴스들을 지정하는 상기 오토 스케일링 그룹 내의 멤버십에 기초하여 선택되는 컴퓨터 구현 방법.

청구항 4

제2항에 있어서, 상기 선택 기준들은 보안 그룹의 멤버들에 적용되는 하나 이상의 네트워크 액세스 규칙을 지정하는 상기 보안 그룹의 멤버들인 가상 머신 인스턴스들을 배제하는 컴퓨터 구현 방법.

청구항 5

제2항에 있어서, 상기 선택된 활성 애플리케이션 구성요소를 종결시키는 단계는 상기 컴퓨팅 클라우드 내에서 선택된 상기 가상 머신 인스턴스의 실행을 종결시키는 단계를 포함하는 컴퓨터 구현 방법.

청구항 6

삭제

청구항 7

제1항에 있어서, 상기 선택된 활성 애플리케이션 구성요소의 종결에 기인하는 상기 나머지 활성 애플리케이션 구성요소들에의 영향을 기록하는 하나 이상의 로그 엔트리를 생성하는 단계를 더 포함하는 컴퓨터 구현 방법.

청구항 8

처리 유닛에 의해 실행될 때, 상기 처리 유닛으로 하여금 네트워크화된 애플리케이션의 리질리언시를 검증하는 동작을 수행하게 하는 명령어들을 포함하는 컴퓨터 관독가능 저장 매체로서, 상기 동작은,

상기 네트워크화된 애플리케이션을 제공하기 위해 사용되는 네트워크 내에서 복수의 활성 애플리케이션 구성요소를 식별하는 단계;

복수의 선택 기준들에 기초하여, 상기 식별된 애플리케이션 구성요소들 중 적어도 하나의 애플리케이션 구성요소를 선택하는 단계;

상기 복수의 선택 기준들에서 특정된 타이밍 정보에 기초하여, 상기 네트워크화된 애플리케이션 내의 상기 선택된 활성 애플리케이션 구성요소의 무작위 고장을 시뮬레이션하기 위하여 상기 선택된 활성 애플리케이션 구성요소를 종결시키는 단계; 및

상기 선택된 활성 애플리케이션 구성요소의 종결 후에, 상기 네트워크 내의 하나 이상의 나머지 활성 애플리케이션 구성요소들을 모니터링하는 단계

를 포함하고,

상기 선택된 활성 애플리케이션 구성요소는 가상 머신 인스턴스 상에서 실행되는 지정된 프로세스를 포함하고, 상기 선택된 애플리케이션 구성요소를 종결시키는 단계는 상기 지정된 프로세스의 실행을 정지시키는 단계를 포함하는 컴퓨터 관독가능 저장 매체.

청구항 9

제8항에 있어서, 상기 복수의 활성 애플리케이션 구성요소 각각은 컴퓨팅 클라우드에서 실행되는 가상 머신 인스턴스를 포함하는 컴퓨터 관독가능 저장 매체.

청구항 10

제9항에 있어서, 상기 가상 머신 인스턴스는 오토 스케일링 그룹 내에서 유발되는 최소 또는 최대 수의 가상 머신 인스턴스들을 지정하는 상기 오토 스케일링 그룹 내의 멤버십에 기초하여 선택되는 컴퓨터 관독가능 저장 매체.

청구항 11

제9항에 있어서, 상기 선택 기준들은 보안 그룹의 멤버들에 적용되는 하나 이상의 네트워크 액세스 규칙들을 지정하는 상기 보안 그룹의 멤버들인 가상 머신 인스턴스들을 배제하는 컴퓨터 관독가능 저장 매체.

청구항 12

제9항에 있어서, 상기 선택된 활성 애플리케이션 구성요소를 종결시키는 단계는 상기 컴퓨팅 클라우드 내에서 선택된 상기 가상 머신 인스턴스의 실행을 종결시키는 단계를 포함하는 컴퓨터 관독가능 저장 매체.

청구항 13

삭제

청구항 14

제8항에 있어서, 상기 동작은 상기 선택된 활성 애플리케이션 구성요소의 종결에 기인하는 상기 나머지 활성 애플리케이션 구성요소들에의 영향을 기록하는 하나 이상의 로그 엔트리들을 생성하는 단계를 더 포함하는 컴퓨터 관독가능 저장 매체.

청구항 15

시스템으로서,

프로세서; 및

상기 프로세서 상에서 실행될 때, 네트워크화된 애플리케이션의 리질리언시를 검증하는 동작을 수행하는 프로그램을 포함하는 메모리

를 포함하고, 상기 동작은,

상기 네트워크화된 애플리케이션을 제공하기 위해 사용되는 네트워크 내에서 복수의 활성 애플리케이션 구성요소를 식별하는 단계;

복수의 선택 기준들에 기초하여, 상기 식별된 애플리케이션 구성요소들 중 적어도 하나의 애플리케이션 구성요소를 선택하는 단계;

상기 복수의 선택 기준들에서 특정된 타이밍 정보에 기초하여, 상기 네트워크화된 애플리케이션 내의 상기 선택된 활성 애플리케이션 구성요소의 무작위 고장을 시뮬레이션하기 위하여 상기 선택된 활성 애플리케이션 구성요소를 종결시키는 단계; 및

상기 선택된 활성 애플리케이션 구성요소의 종결 후에, 상기 네트워크 내의 하나 이상의 나머지 활성 애플리케이션 구성요소들을 모니터링하는 단계

를 포함하고,

상기 선택된 활성 애플리케이션 구성요소는 가상 머신 인스턴스 상에서 실행되는 지정된 프로세스를 포함하고, 상기 선택된 애플리케이션 구성요소를 종결시키는 단계는 상기 지정된 프로세스의 실행을 정지시키는 단계를 포함하는 시스템.

청구항 16

제15항에 있어서, 상기 복수의 활성 애플리케이션 구성요소 각각은 컴퓨팅 클라우드에서 실행되는 가상 머신 인스턴스를 포함하는 시스템.

청구항 17

제16항에 있어서, 상기 가상 머신 인스턴스는 오토 스케일링 그룹 내에서 유발되는 최소 또는 최대 수의 가상 머신 인스턴스들을 지정하는 상기 오토 스케일링 그룹 내의 멤버십에 기초하여 선택되는 시스템.

청구항 18

제16항에 있어서, 상기 선택 기준들은 보안 그룹의 멤버들에 적용되는 하나 이상의 네트워크 액세스 규칙들을 지정하는 상기 보안 그룹의 멤버들인 가상 머신 인스턴스들을 배제하는 시스템.

청구항 19

제16항에 있어서, 상기 선택된 활성 애플리케이션 구성요소를 종결시키는 단계는 상기 컴퓨팅 클라우드 내에서 선택된 상기 가상 머신 인스턴스의 실행을 종결시키는 단계를 포함하는 시스템.

청구항 20

삭제

청구항 21

제15항에 있어서, 상기 동작은 상기 선택된 활성 애플리케이션 구성요소의 종결에 기인하는 상기 나머지 활성 애플리케이션 구성요소들에의 영향을 기록하는 하나 이상의 로그 엔트리들을 생성하는 단계를 더 포함하는 시스템.

발명의 설명

기술 분야

<관련 출원들에 대한 상호 참조>

본 출원은 본 명세서에 참조문헌으로 포함되어 있는, 2010년 9월 20일자로 출원된 미국 특허 출원 일련번호 제 12/886,138호의 이익을 주장한다.

<기술분야>

[0004] 본 발명의 실시예들은 일반적으로 상호작용하는 서버들의 컬렉션에 의해 제공되는 분산 컴퓨팅 서비스의 리질리언시(resiliency)를 평가하는 기술들에 관한 것이다.

배경기술

[0005] 다양한 컴퓨팅 애플리케이션들이 컴퓨터 네트워크들을 통해 사용자들에게 이용가능해졌다. 종종, 네트워크화된 애플리케이션은 다수의 상호작용하는 컴퓨팅 서버들을 사용하여 제공될 수 있다. 예를 들어, 웹 사이트는 웹 페이지들에 대한 사용자들로부터의 요청들을 수신하도록 구성된 웹 서버(한 컴퓨팅 시스템 상에서 실행됨)를 사용하여 제공될 수 있다. 요청들이 애플리케이션 서버(다른 컴퓨팅 시스템 상에서 실행됨)에 전달될 수 있고, 애플리케이션 서버가 결국 요청들을 처리하고 웹 서버, 및 궁극적으로는 사용자들에게 다시 통과되는 응답들을 생성한다.

[0006] 다른 예는 네트워크를 통해 미디어 타이틀에의 액세스를 제공하기 위해 사용되는 콘텐츠 배포 시스템을 포함한다. 전형적으로, 콘텐츠 배포 시스템은 클라이언트들이 게임 콘솔, 컴퓨팅 시스템, 컴퓨팅 태블릿, 이동 전화, 네트워크 인식 DVD 플레이어 등과 같은 콘텐츠 플레이어를 사용하여 접속하는 액세스 서버들, 콘텐츠 서버들 등을 포함할 수 있다. 콘텐츠 서버는 콘텐츠 서버로부터 콘텐츠 플레이어로의 다운로드를 위해 이용가능한 파일들(또는 "스트림들")을 저장한다. 각 스트림은 디지털 버전의 영화, 텔레비전 프로그램, 스포츠 이벤트, 사용자 생성된 콘텐츠, 기록된 비디오에 의해 캡처되는 연출 또는 라이브 이벤트 등을 제공할 수 있다. 사용자들은 콘텐츠의 리스트가 이용가능한 웹 서버에 연결함으로써 서비스에 액세스한다. 특정 타이틀에 대한 요청이 수신되면, 그것은 이용가능한 콘텐츠 서버로의 접속을 통해 클라이언트 시스템에 스트리밍될 수 있다.

[0007] 이들과 같은 시스템들 상에서 실행되는 소프트웨어 애플리케이션들은 종종 취약성을 및 에러들을 조정하기 위한 패치들은 물론, 새로운 특징들을 이용가능하게 하기 위한 업그레이들에 있어서의 진행 중인 개발 결과로서 업데이트된다. 동시에, 네트워크화된 애플리케이션에서의 서버들은 예측하지 못하거나 의도되지 않은 방식으로 서로 의존할 수 있고, 한 시스템에 대한 변경들은 다른 시스템에 대한 의도되지 않은 종속성을 야기할 수 있다. 이러한 일이 발생한 때, 서버가 고장나면, 네트워크화된 애플리케이션에 대한 액세스가 중단될 수 있다.

발명의 내용

[0008] 본 명세서에서 개시된 본 발명의 일 실시예는 네트워크화된 애플리케이션의 리질리언시를 검증하는 컴퓨터 구현 방법을 제공한다. 방법은 일반적으로 네트워크화된 애플리케이션을 제공하기 위해 사용되는 네트워크 내에서 복수의 활성 애플리케이션 구성요소들을 식별하는 단계, 및 하나 이상의 선택 기준들에 기초하여, 식별된 애플리케이션 구성요소들 중 적어도 하나를 선택하는 단계를 포함할 수 있다. 이 방법은 선택된 활성 애플리케이션 구성요소를 종결시키는 단계, 및 선택된 활성 애플리케이션 구성요소의 종결 후에, 네트워크 내의 하나 이상의 나머지 활성 애플리케이션 구성요소들을 모니터링하는 단계도 포함할 수 있다.

[0009] 다른 실시예들은 처리 유닛으로 하여금 개시된 방법들의 하나 이상의 양상들을 구현할 수 있게 하는 명령어들을 포함하는 컴퓨터 관독가능 매체뿐만 아니라 개시된 방법들의 하나 이상의 양상들을 구현하도록 구성된 시스템을 제한 없이 포함한다.

도면의 간단한 설명

[0010] 위에 열거한 본 발명의 특징들이 상세히 이해될 수 있도록, 위에 간단히 요약된 본 발명의 더 상세한 설명이 실시예들을 참조하여 행해질 수 있으며, 실시예들 중 일부는 첨부 도면들에 예시된다. 그러나, 본 발명은 다른 동등한 효과의 실시예들을 인정할 수 있으므로, 첨부 도면들은 본 발명의 전형적인 실시예들을 예시할 뿐이며, 따라서 그 범위의 제한으로 간주되어서는 안 된다는 것이 주목되어야 한다.

도 1은 본 발명의 하나 이상의 양상들을 구현하도록 구성된 컴퓨팅 인프라스트럭처를 예시한다.

도 2는 본 발명의 일 실시예에 따른 클라우드 컴퓨팅 환경 내의 복수의 상호작용하는 서버 인스턴스들을 예시한다.

도 3은 본 발명의 일 실시예에 따른 리질리언시 모니터링 애플리케이션을 포함하는 컴퓨팅 시스템의 도면이다.

도 4는 본 발명의 일 실시예에 따른 네트워크화된 애플리케이션들의 리질리언시를 검증하는 방법을 예시한다.

발명을 실시하기 위한 구체적인 내용

[0011] 본 발명의 실시예들은 상호작용하는 서버들의 컬렉션을 사용하여 이용가능해진 네트워크화된 애플리케이션의 리질리언시를 검증하는 기술들을 제공한다. 예를 들어, 네트워크 모니터링 애플리케이션은 고장난 것에 의존하는 시스템들이 무작위의 예기치 않은 고장 후에도 여전히 올바르게 기능(또는 기품 있게 퇴화(degrade gracefully))할 수 있는지를 판단하기 위해, 실행 중인 애플리케이션의 인스턴스를 종결시키도록 구성될 수 있다. 따라서, 모니터링 애플리케이션은 네트워크화된 애플리케이션 내의 다른 시스템들에 대한 서버 고장의 영향을 제어된 방식으로 관찰할 수 있다. 이 접근법은 임의의 서버가 임의의 시간에 사라질 수 있는 클라우드 기반 전개들에 유용할 수 있다.

[0012] 일 실시예에서, 네트워크 모니터링 애플리케이션은 볼특정 간격들에서 각 실행 서버(또는 애플리케이션)를 관찰하고, 하나를 선택하여 종결시킨다. 클라우드 기반 전개의 경우에, 이것은 가상 머신 인스턴스를 종결시키는 것, 서버 상에서 실행되는 프로세스를 종결시키는 것 등을 포함할 수 있다. 데이터 센터 내의 물리적 서버들에 대해서는, 서버를 정지시키는 것, 서버 상에서 실행되는 프로세스를 종결시키는 것, 서버 상에서 네트워크 접속을 폐쇄하는 것 등을 포함할 수 있다. 어떻게 수행되더라도, 네트워크 애플리케이션에서의 선택된 서버의 참여가 종료되어, 나머지 네트워크 애플리케이션으로부터 그 서버(또는 애플리케이션)를 차단시킨다.

[0013] 나머지 네트워크 애플리케이션에 대한 고장난 서버의 효과들을 관찰함으로써, 제공자는 각 구성요소가 경고 없이 사라지는 임의의 단일 인스턴스를 용인할 수 있는 것을 보장할 수 있다. 일 실시예에서, 네트워크 모니터링 애플리케이션은 업데이트 또는 폐치를 제작 환경 내의 서버들(또는 애플리케이션들)에 전개하기 전에 테스트 환경에서 사용될 수 있다. 그렇게 하는 것은 업데이트 또는 폐치의 효과들이 제작 환경에 전개되는 것 없이 평가되는 것을 가능하게 한다. 게다가, 어떤 애플리케이션들(또는 호스트들 또는 시스템들)은 배제/포함 리스트를 사용하여, 가능한 종결로부터 배제(또는 포함)될 수 있다. 유사하게, 클라우드 기반 전개에서, 네트워크 모니터링 애플리케이션은 오토 스케일링 그룹(auto scaling group)의 멤버들인 서버 인스턴스들을 종결시키도록 구성될 수 있다. 그렇게 하는 것은 오토 스케일링 프로세스들의 기능이, 무작위로 발생하는 서버 고장들에 대해 평가되는 것을 가능하게 한다. 따라서, 다양한 실시예들에서, 네트워크 모니터링 애플리케이션은 의도하지 않거나 원하지 않는 종속성들을 생성하는 방식들로 제작 시스템들이 업그레이드, 폐치, 또는 다르게 변경될 때 시간에 따라 상실될 수 있었을 내고장성에 대한 요건들을 실시하는 데 도움이 된다. 더 일반적으로, 시스템들의 임의의 논리 그룹은 본 명세서에서 설명되는 네트워크 모니터링 애플리케이션에 의해 정의 및 테스트될 수 있다.

[0014] 이하의 설명에서, 다수의 구체적인 상세들은 본 발명의 더 완전한 이해를 제공하기 위해 제시된다. 그러나, 본 발명은 이를 구체적인 상세들 중 하나 이상이 없더라도 실시될 수 있는 것이 당업자에게 분명할 것이다. 다른 경우들에서, 공지된 특징들은 본 발명의 모호함을 회피하기 위해 설명되지 않았다.

[0015] 게다가, 본 발명의 특정 실시예들은 영화들, 음악, 텔레비전 프로그래밍, 사용자 생성 콘텐츠 등을 데이터 통신 네트워크를 통해 최종 사용자 클라이언트 장치들로 스트리밍하기 위해 사용되는 네트워크화된 애플리케이션의 예를 사용하여 설명된다. 그러나, 본 발명의 실시예들은 다양한 네트워크화된 애플리케이션들 또는 서비스들을 위해 개별 시스템 고장에 대한 리질리언시를 검증하도록 적응될 수 있는 것이 이해되어야 한다. 따라서, 스트리밍 미디어 서비스에 대한 참조들은 예시적일 뿐이고 한정되는 것이 아니다.

[0016] 도 1은 본 발명의 하나 이상의 양상들을 구현하도록 구성된 컴퓨팅 인프라스트럭처(100)를 예시한다. 이 예에서, 컴퓨팅 인프라스트럭처(100)는 네트워크화된 애플리케이션 또는 서비스를 클라이언트 시스템들에 제공하기 위해 사용되는 인프라스트럭처를 나타낸다. 도시된 바와 같이, 데이터 센터(130)에서의 서버 시스템들(137) 및 클라우드 제공자(140)에서 실행되는 가상 머신 인스턴스들(145)은 네트워크 애플리케이션 또는 서비스를 네트워크(120)를 통해 클라이언트들(105₁₋₄)에 제공하기 위해 사용된다. 부가적으로, 데이터 센터(130)는 가용성/신뢰성 모니터링 시스템(135)을 포함한다.

[0017] 언급된 바와 같이, 본 발명의 실시예들은 클라우드 제공자(140)에 의해 이용가능해진 클라우드 컴퓨팅 인프라스트럭처를 사용하여 전개되는 가상 머신 인스턴스들(145)의 리질리언시를 검증하기 위해 사용될 수 있다. 클라우드 컴퓨팅은 일반적으로 네트워크를 통한 서비스로서의 스케일링 가능한 컴퓨팅 자원들의 프로비저닝(provision)을 지칭한다. 더 공식적으로, 클라우드 컴퓨팅은 컴퓨팅 자원과 그의 근본적인 기술적 아키텍처(예를 들어, 서버들, 스토리지, 네트워크들) 사이에 추상화를 제공하여, 최소의 관리 노력 또는 서비스 제공자 상호작용으로 신속히 프로비저닝되어 릴리즈될 수 있는 구성 가능한 컴퓨팅 자원들의 폴(pool)에의 편리한 주문형 네트워크 액세스를 가능하게 하는 컴퓨팅 능력으로서 정의될 수 있다. 따라서, 클라우드 컴퓨팅은, 사용자가 컴퓨팅 자원들을 제공하기 위해 사용되는 근본적인 물리적 시스템들(또는 그 시스템들의 위치들)을 고려하지 않

고서도 "클라우드" 내의 가상 컴퓨팅 자원들(예를 들어, 스토리지, 데이터, 애플리케이션들, 및 심지어 완전한 가상화 컴퓨팅 시스템들)에 액세스하는 것을 가능하게 한다.

[0018] 전형적으로, 클라우드 컴퓨팅 자원들은 사용량별 요금제(pay-per-use basis)로 사용자에게 제공되며, 사용자들은 실제 사용된 컴퓨팅 자원들(예를 들어, 사용자에 의해 소비되는 저장 공간의 양 또는 사용자에 의해 유발된(spawned) 가상 머신 인스턴스들의 수)에 대해서만 요금을 청구받는다. 사용자는 언제 어디에서든지 클라우드에 상주하는 자원들 중 임의의 것에 액세스할 수 있다. 프로비저닝되면, 가상 머신 인스턴스(145)는 컴퓨팅 서버의 추상화를 제공하고, 사용자는 일반적으로 가상 머신 인스턴스(145)에 의해 표현되는 물리적 컴퓨팅 서버를 제어한다는 사상과 동일한 방식으로 가상 머신 인스턴스(145) 상에 애플리케이션들을 설치 및 실행할 수 있다.

[0019] 본 발명의 맥락에서, 서비스 제공자는 가상 머신 인스턴스(145) 상에 서버들 또는 애플리케이션들을 전개하고, 클라이언트들(105)이 데이터 센터(130) 내의 물리적 서버 시스템들(137)에 액세스하는 것과 동일한 방식으로 애플리케이션들(예를 들어, 스트리밍 미디어 서비스)에 접속 및 액세스하는 것을 가능하게 할 수 있다. 동시에, 서비스 제공자는 단순히 추가 가상 머신 인스턴스들(145)을 유발시킴으로써 서비스를 신속히 스케일링할 수 있다. 이것은 대형 컴퓨팅 인프라스트럭처를 데이터 센터(130)에 구축 및 유지할 필요 없이 제공자가 퍼크 요구 기간들에 반응하는 것을 가능하게 한다. 클라우드 제공자(140)는 임의의 주어진 시간에 주어진 애플리케이션(또는 애플리케이션 구성요소)의 요구들에 기초하여 그것에 할당된 가상 머신 인스턴스들(145)의 수를 자동으로 확대 또는 축소하기 위해 사용되는 오토 스케일링 특징을 제공할 수 있다.

[0020] 그러나, 서비스 제공자가 근본적인 컴퓨팅 하드웨어를 제어하지 않으므로, 가상 머신 인스턴스들(145) 상에서 실행되는 서버들 또는 애플리케이션들은 경고 없이 사라지는 임의의 단일 가상 인스턴스(또는 가상 머신 인스턴스에 의해 제공되는 서비스)를 용인하도록 구성되어야 한다. 따라서, 아래에 더 상세히 설명되는 바와 같이, 모니터링 시스템(135)은 서버 시스템들(137) 또는 가상 머신 인스턴스들(145) 상에서 실행 애플리케이션의 인스턴스를 주기적으로 종결시키고 (제작 또는 테스트 환경들에서) 서비스 전체에 대한 영향을 관찰하도록 구성된 애플리케이션 프로그램을 포함할 수 있다. 그리고, 가상 머신 인스턴스의 종결 후에, 오토 스케일링 그룹과 관련된 인스턴스들이 적절히 스케일링되는 것을 보장할 수도 있다. 알려지지 않거나 원하지 않는 의존성들이 식별되는 경우에, 애플리케이션들은 적절히 리팩토링될(refactored) 수 있다.

[0021] 도 2는 본 발명의 일 실시예에 따른 컴퓨팅 클라우드(200) 내의 복수의 상호작용하는 서버 인스턴스들을 예시한다. 예시적으로, 컴퓨팅 클라우드(200)는 스트리밍 미디어 서비스를 클라이언트 시스템들에 제공하기 위해 사용되는 컴퓨팅 인프라스트럭처의 예를 제공한다. 물론, 상기 언급된 바와 같이, 컴퓨팅 클라우드(200)는 다양한 컴퓨팅 서비스들을 클라이언트들(105)에 제공하기 위해 사용될 수 있다.

[0022] 도시된 바와 같이, 컴퓨팅 클라우드(200)는 4개의 상이한 오토 스케일링 그룹들(205₁₋₄) 간에 할당된 가상 머신 인스턴스들(220, 225, 230 및 235)을 포함한다. 각 오토 스케일링 그룹은 서비스들에 대한 요구에 기초하여, 프로비저닝되어야 하는 최소 및/또는 최대 수의 인스턴스들과 관련될 수 있다. 이 예에서, 제1 오토 스케일링 그룹(205₁)은 클라이언트 시스템들(105)로부터 초기 요청들을 수신하기 위해 사용되는 웹 서버 인스턴스들(220)을 포함한다. 요청이 수신되고 나면, 그 요청은 제2 오토 스케일링 그룹(205₄) 내의 애플리케이션 서버 인스턴스들(235) 중 하나에 전달되는데, 이는 웹 서버 인스턴스(220)에 다시 전달되는 웹 페이지에 대한 콘텐츠를 생성할 수 있으며, 거기에서 그것은 요청 클라이언트 시스템(105)에 서빙된다. 예를 들어, 초기 웹 페이지는 사용자가 스트리밍 미디어 콘텐츠에 액세스하기 위해 자격증명서들을 제출하는 것을 허용하는 양식(form)을 포함할 수 있다. 그러한 경우에, 자격증명서들은 웹 서버 인스턴스(220) 및 애플리케이션 서버 인스턴스(235)에 다시 전달된다. 그리고 다음으로, 애플리케이션 서버(235)는 제3 오토 스케일링 그룹(205₃) 내의 데이터베이스 서버 인스턴스들(230) 중 하나와 통신함으로써 자격증명서들을 검증할 수 있다. 예를 들어, 데이터베이스 서버 인스턴스(230)는 자격증명서들을 사용하여 결정된 사용자에 대한 가입 상태를 나타내는 정보를 데이터베이스(210)로부터 검색할 수 있다. 인증되면, 애플리케이션 서버 인스턴스(235)는 클라이언트(105) 또는 웹 서버 인스턴스(220)에 전달되는 스트리밍을 위해 이용가능한 미디어 타이틀들을 보여주는 웹 페이지들을 생성한다.

[0023] 그 후, 클라이언트가 타이틀을 스트리밍하라고 요청할 때, (제4 오토 스케일링 그룹(205₂) 내의) 콘텐츠 스트리밍 인스턴스들(225) 중 하나는 콘텐츠 데이터베이스(215)로부터 스트리밍 미디어 데이터를 검색하고 그것을 요청 클라이언트 시스템(105)에 송신한다. 스트리밍 미디어 서비스가 제공자의 데이터 센터로부터 호스트되는 경우에, 가상 머신 인스턴스들(220, 225, 230 및 235)은 일반적으로 데이터 센터 내의 물리적 컴퓨팅 시스템들에 대응한다.

[0024]

일 실시예에서, 모니터링 시스템(135)은 컴퓨팅 클라우드(200)(또는 제공자의 데이터 센터 내의 시스템들/애플리케이션들)에 의해 제공되는 스트리밍 미디어 서비스의 리질리언시를 평가하도록 구성될 수 있다. 예를 들어, 모니터링 시스템(135)은 인스턴스들(220, 225, 230 또는 235) 중 하나(또는 오토 스케일링 그룹들(205-4) 중 지정된 것으로부터 선택된 인스턴스)를 종결시키는 것을 선택할 수 있다. 선택은 무작위 간격들로 수행될 수 있거나 예정대로 발생할 수 있다. 인스턴스를 종결시키는 것은 종결된 것에 의존하는 시스템들이 무작위의 예기치 않은 고장 후에 계속해서 올바르게 기능(또는 기품 있게 퇴화)하는지를 제공자가 평가하는 것을 가능하게 한다. 예를 들어, 애플리케이션 서버 인스턴스들(235) 중 하나가 종결되면, 콘텐츠를 클라이언트들(105)에 계속해서 스트리밍하는 콘텐츠 스트리밍 인스턴스들(225)의 능력이 관찰될 수 있다. 따라서, 모니터링 시스템(135)은 사용자들이 네트워크화된 애플리케이션 내의 다른 시스템들에 대한 서버 고장의 영향을 제어된 방식으로 관찰하는 것을 가능하게 한다

[0025]

일 실시예에서, 인스턴스들(220, 225, 230 및 235) 중 일부는 모니터링 시스템(135)에 의해 종결될 자격이 있는 것으로부터 배제될 수 있다. 개별 인스턴스들은 배제 리스트를 사용하여 배제될 수 있다. 그러한 리스트는 개별 인스턴스들을 인스턴스 ID를 사용하여 종결할 자격이 있는 것으로부터 면제할 수 있다. 예를 들어, 인증 서버로서 사용되는 인스턴스는 종결할 자격으로부터 배제될 수 있다. 유사하게, 하나의 인스턴스(또는 그 인스턴스에 의해 제공되는 애플리케이션)의 고장이 다른 것들에 지장이 있는 것으로 알려지면, 그것은 종결할 자격으로부터 배제될 수 있다.

[0026]

부가적으로, 인스턴스들은 그룹 멤버십에 기초하여 면제될 수 있다. 예를 들어, 모니터링 시스템(135)은 지정된 오토 스케일링 그룹 내의 모든 인스턴스들을 배제하도록 구성될 수 있다. 다른 예는 특정 보안 그룹에 속하는 모든 인스턴스들을 배제할 것이다. 이 맥락에서, 보안 그룹은 그룹의 멤버인 임의의 인스턴스에 방화벽과 같은 액세스 규칙들의 그룹이 적용되는 시스템들의 그룹인 것에 주목한다. 예를 들어, 데이터베이스 서버 인스턴스들(230)은 애플리케이션 서버 인스턴스들(235)에 대한 액세스를 허용하지만, 웹 서버 인스턴스들(220)에 대한 액세스를 차단하는 "데이터베이스" 그룹의 멤버일 수 있다. 유사하게, 웹 서버 인스턴스들(220)은 "웹 그룹"에 속하고, 지정된 포트(예를 들어, HTTP 트래픽에 대한 포트 80) 상의 공중 인터넷에의 액세스를 승인받을 수 있다. 물론, 상기 논의된 오토 스케일링 및 보안 그룹 예들 외에도, 시스템들의 다른 논리적 그룹들이 네트워크 모니터링 애플리케이션에 의해 정의 및 테스트될 수 있다.

[0027]

컴퓨팅 클라우드(200)의 외부에 도시되어 있긴 하지만, 모니터링 시스템(135)은 컴퓨팅 클라우드(200)에서 유발된 가상 머신 인스턴스를 자체 실행할 수 있는 것에 주목한다.

[0028]

도 3은 본 발명의 일 실시예에 따른 리질리언시 모니터링 애플리케이션을 포함하는 모니터링 시스템(135)의 도면이다. 도시된 바와 같이, 모니터링 시스템(135)은 제한 없이, CPU(central processing unit)(205), 네트워크 인터페이스(315), 상호접속(320), 메모리(325), 및 스토리지(330)를 포함한다. 모니터링 시스템(135)은 I/O 장치들(212)(예를 들어, 키보드, 디스플레이 및 마우스 장치들)을 모니터링 시스템(135)에 접속하는 I/O 장치 인터페이스(310)도 포함할 수 있다.

[0029]

일반적으로, CPU(305)는 메모리(325)에 저장되는 프로그래밍 명령어들을 검색 및 실행한다. 유사하게, CPU(305)는 메모리(325)에 상주하는 애플리케이션 데이터를 저장 및 검색한다. 상호접속(320)은 CPU(305), I/O 장치 인터페이스(310), 스토리지(330), 네트워크 인터페이스(315), 및 메모리(325) 사이에서 프로그래밍 명령어들 및 애플리케이션 데이터의 송신을 용이하게 한다. CPU(305)는 단일 CPU, 다수의 CPU들, 다수의 처리 코어들을 갖는 단일 CPU 등을 나타내기 위해 포함된다. 그리고, 메모리(325)는 일반적으로 랜덤 액세스 메모리를 나타내기 위해 포함된다. 스토리지(330)는 디스크 드라이브 저장 장치일 수 있다. 단일 유닛으로서 도시되어 있지만, 스토리지(330)는 고정 디스크 드라이브들, 플로피 디스크 드라이브들, 테이프 드라이브들, 이동식 메모리 카드들, 광 스토리지, NAS(network attached storage), 또는 SAN(storage area-network)과 같은 고정 및/또는 이동식 저장 장치들의 조합일 수 있다.

[0030]

예시적으로, 메모리(325)는 모니터링 애플리케이션(321)을 포함하고 스토리지(330)는 모니터링 로그들(335)을 포함한다. 도시된 바와 같이, 모니터링 애플리케이션(321)은 종결 구성요소(323), 복구 모니터(327), 및 인스턴스 모니터링 파라미터들(329)을 포함한다. 상기 언급된 바와 같이, 모니터링 애플리케이션(321)은 네트워크화된 애플리케이션(예를 들어, 클라우드 컴퓨팅 클라우드(200) 내의 가상 머신 인스턴스들(220, 225, 230 또는 235) 중 하나 또는 제공자의 데이터 센터 내의 서버 상에서 실행하는 애플리케이션)에 사용되는 실행 중인 가상 머신 인스턴스, 서버, 애플리케이션 또는 다른 구성요소를 주기적으로 선택 및 종결시키도록 구성된 소프트웨어 애플리케이션을 제공할 수 있다. 편의상, "인스턴스들"의 종결이 참조될 것이지만, 임의의 다른 구성요소의 종

결이 예상되는 것이 이해될 것이다.

[0031] 일 실시예에서, 종결 구성요소(323)는 모니터링 파라미터들(329)에 따라 어느 인스턴스를 종결시킬지(뿐만 아니라 인스턴스를 종결시킬 시기)를 선택한다. 선택 파라미터들은 배제된 인스턴스들 또는 인스턴스들의 그룹들, 시각들(times of day) 등과 같은 기준들을 지정할 수 있는데, 종결 구성요소(323)는 종결 선택을 위해 이를 사용할 수 있다. 예를 들어, 모니터링 파라미터들(329)은 임의의 시간에 또는 지정된 시간 간격 동안 인스턴스를 무작위로 선택(또는 그룹 중에서 무작위로 선택)할 것을 지정할 수 있다.

[0032] 인스턴스가 선택(및 종결)되면, 복구 모니터(327)는 네트워크화된 애플리케이션의 나머지 인스턴스들의 작용들을 관찰하고, 대응하는 정보를 생성할 수 있으며, 복구 모니터(327)는 다음으로 이 정보를 로그들(335)에 저장한다. 로그들(335)의 콘텐츠는 모니터링 파라미터들(329)에 의해 지정된 정보를 포함할 뿐만 아니라 인스턴스들(또는 인스턴스 상에서 실행하는 애플리케이션들)에 의해 네이티브로 생성된 로깅 데이터를 포함할 수 있다. 즉, 인스턴스 상에서 실행하는 애플리케이션들은 그 애플리케이션들이 거기에서 실행되는 것에 따라 로그들(예를 들어, 웹 서버에 대한 액세스 이력 로그)을 생성할 수 있다.

[0033] 도 4는 본 발명의 일 실시예에 따른 네트워크화된 애플리케이션들의 리질리언시를 검증하는 방법(400)을 예시한다. 도시된 바와 같이, 방법(400)은 모니터링 파라미터들에 의해 지정된 구성에 따라 모니터링 애플리케이션의 초기화되는 단계(405)에서 시작한다. 예를 들어, 구성 파라미터들은 종결시킬 인스턴스를 선택할 시기를 판단할 뿐만 아니라, 어느 인스턴스, 애플리케이션, 또는 서버를 종결시킬지를 선택하는 기준들을 지정할 수 있다. 단계(410)에서, 모니터링 애플리케이션은 실행 중인 인스턴스를 종결시킬 시간에 도달할 때까지 대기한다. 도달되면, 단계(415)에서, 모니터링 애플리케이션은 네트워크화된 애플리케이션을 제공하기 위해 복수의 활성 애플리케이션 구성요소들(예를 들어, 활성 가상 머신 인스턴스들, 애플리케이션들, 또는 프로세스들)이 사용되는 것을 식별한다.

[0034] 단계(420)에서, 모니터링 애플리케이션은 종결될 가상 머신 인스턴스(또는 프로세스 또는 애플리케이션)를 선택한다. 선택되면, 모니터링 애플리케이션은 메시지를 송신하여 선택된 인스턴스를 종료(kill)시킨다. 예를 들어, 가상 머신 인스턴스의 경우에, 모니터링 애플리케이션은 종결 인스턴스 메시지를 클라우드 네트워크에 송신할 수 있다. 대안적으로, 모니터링 애플리케이션은 가상 머신 인스턴스 상의(또는 데이터 센터 서버 상의) 서버 프로그램(예를 들어, HTTP 웹 서버)을 셧다운시키거나 운영 체제 상에 제공된 메커니즘들을 사용하여 프로세스를 종료시킬 수 있다.

[0035] 단계(425)에서, 모니터링 애플리케이션은 선택된 인스턴스가 종결되는(또는 그렇지 않으면 셧다운되거나 실행 중지됨) 것을 대기한다. 선택된 인스턴스의 종결 후에, 단계(430)에서, 모니터링 애플리케이션은 나머지 인스턴스들(또는 애플리케이션들)의 동작을 관찰하고 로그 데이터를 기록하여 종결된 인스턴스의 소멸이 나머지 네트워크화된 애플리케이션에 어떻게 영향을 주는지를 캡처한다. 나머지 네트워크화된 애플리케이션 상에서 종결된 것의 효과들을 관찰함으로써, 제공자는 각 구성요소가 경고 없이 사라지는 임의의 단일 인스턴스를 용인할 수 있는 것을 보장할 수 있다.

[0036] 유리하게, 본 발명의 실시예들은 상호작용 서버들의 수집을 사용하여 이용가능해진 네트워크화된 애플리케이션의 리질리언시를 검증하는 기술들을 제공한다. 일 실시예에서, 네트워크 모니터링 애플리케이션은 불특정 간격들에서 각 실행 중인 서버(또는 애플리케이션)를 관찰하고, 하나를 선택하여 종결시킨다. 클라우드 기반 전개의 경우에, 이것은 가상 머신 인스턴스를 종결시키는 것 또는 서버 상에서 실행하는 프로세스를 종결시키는 것을 포함할 수 있다. 그렇게 하는 것은 클라우드 서비스 제공자에 의해 이용가능해진 오토 스케일링(또는 시스템들의 다른 논리적 그룹)의 유효성을 테스트할 수 있다. 부가적으로, 일부 시스템들(또는 시스템들의 그룹들)은 네트워크 모니터링 애플리케이션(예를 들어, 보안 그룹에 속하는 시스템들)에 의해 종결할 자격이 있는 것으로부터 배제될 수 있다.

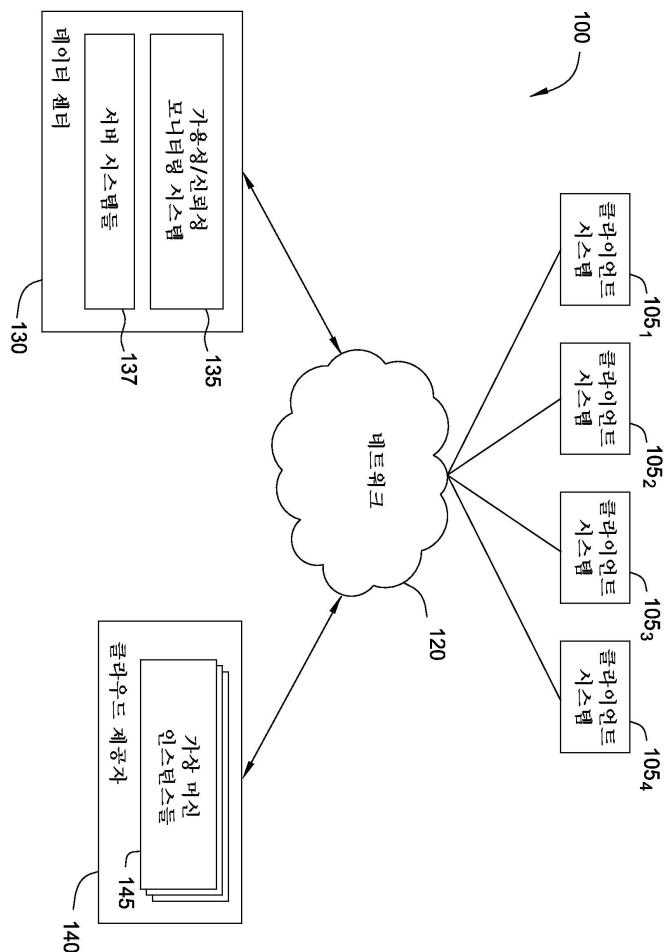
[0037] 전술한 것이 본 발명의 실시예들에 관한 것일지라도, 본 발명의 다른 및 추가 실시예들은 그 기본 범위로부터 벗어나지 않고 안출될 수 있다. 예를 들어, 본 발명의 양상들은 하드웨어 또는 소프트웨어로, 또는 하드웨어 및 소프트웨어의 조합으로 구현될 수 있다. 본 발명의 일 실시예는 컴퓨터 시스템과 함께 사용하기 위한 프로그램 제품으로서 구현될 수 있다. 프로그램 제품의 프로그램(들)은 실시예들(본 명세서에서 설명되는 방법들을 포함함)의 기능들을 정의하며, 다양한 컴퓨터 관독 가능 저장 매체 상에 포함될 수 있다. 예시적인 컴퓨터 관독 가능 저장 매체는 (i) 정보가 영구적으로 저장되는 기록불가능 저장 매체(예를 들어, CD-ROM 드라이브에 의해 관독가능한 CD-ROM 디스크들, 플래시 메모리, ROM 칩들 또는 임의의 타입의 고체 상태 비휘발성 반도체 메모리와 같은 컴퓨터 내의 관독 전용 메모리 장치들); 및 (ii) 변경가능한 정보가 저장되는 기록가능 저장 매체(예를

들어, 디스크 드라이브 또는 하드 디스크 드라이브 내의 플로피 디스크들 또는 임의의 타입의 고체 상태 랜덤 액세스 반도체 메모리)를 포함하지만, 이들에 한정되지 않는다. 그러한 컴퓨터 관독가능 저장 매체는 본 발명의 기능들을 명령하는 컴퓨터 관독가능 명령어들을 지닐 때, 본 발명의 실시예들이다.

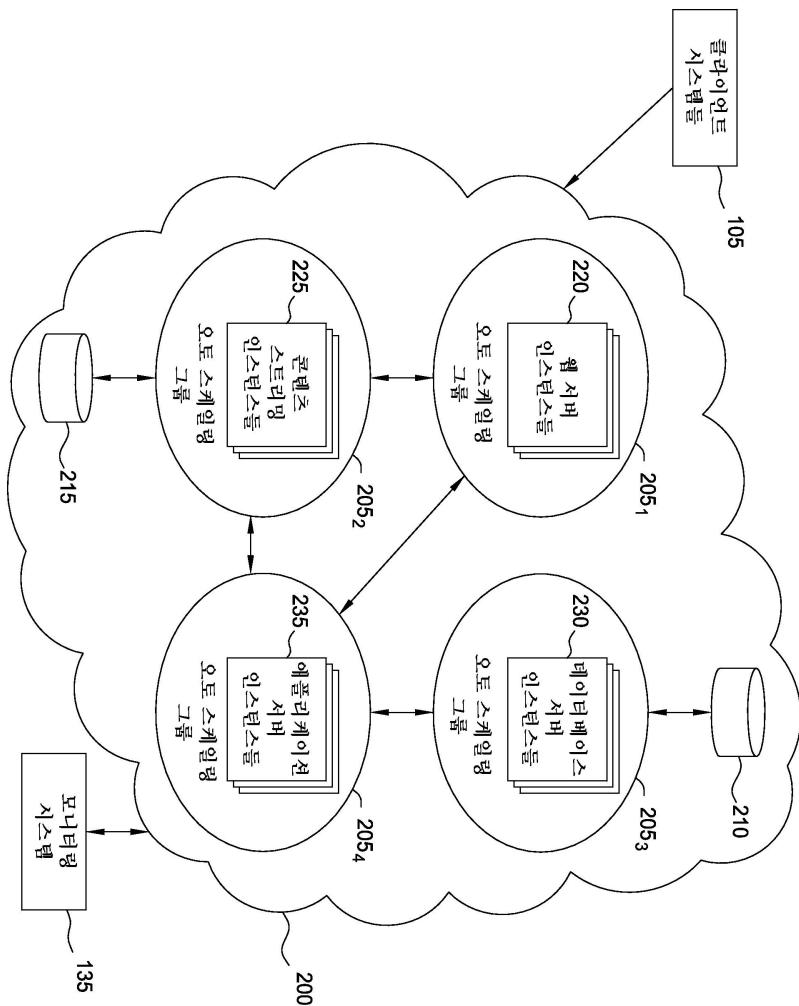
[0038] 그러므로, 본 발명의 범위는 이하의 청구항들에 의해 결정된다.

도면

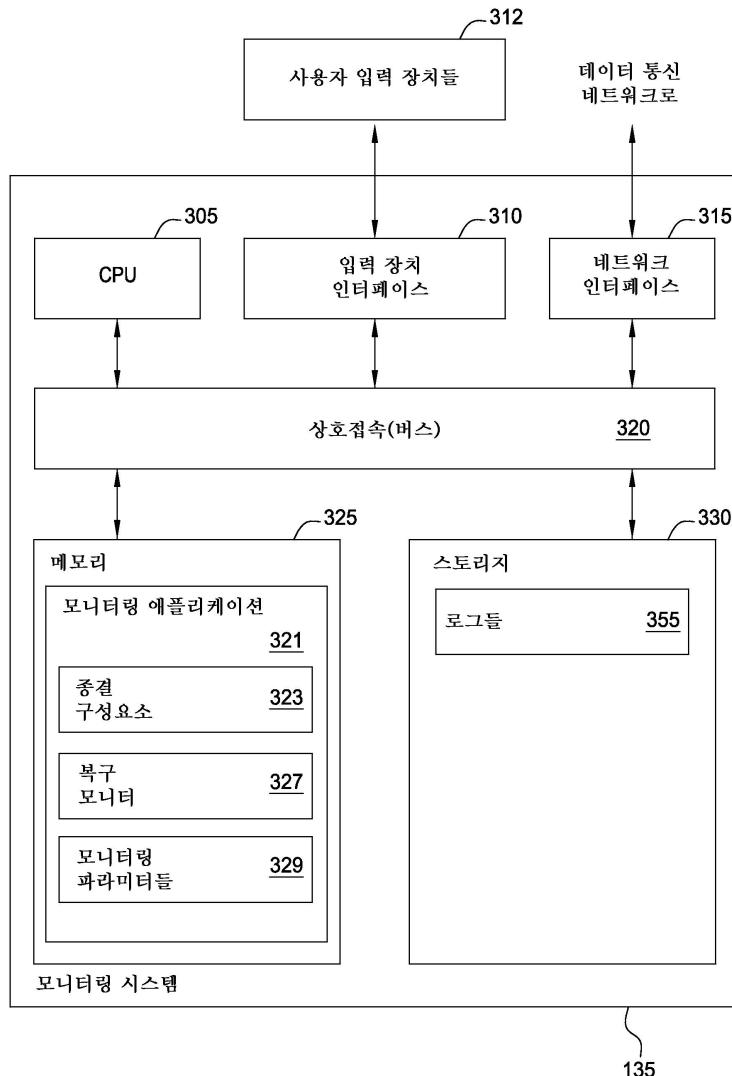
도면1



도면2



도면3



도면4

